Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 9, September 2016, pg.72 – 75

Improving Resource Management & Solving Scheduling Problem in Data Ware House using OLAP & OLTP: A Review

Rekha
M.Tech Student, Department of Computer Science & Applications, Maharishi Dayanand University
rekhapanchal9124@gmail.com

Dr. Rajender Singh Chhillar
Professor & Former Head, Department of Computer Science & Applications, Maharishi Dayanand University
Chhillar02@gmail.com

Abstract

A data warehouse [1] provides us generalized & consolidated data in multidimensional view. Along with generalized & consolidate view of data, a data warehouses also given us Online Analytical Processing (OLAP)[2] tools. These tools help us in interactive & effective analysis of data in a multidimensional space. This analysis results in data generalization & data mining. Data mining functions such as clustering, classification, prediction could be integrated with OLAP operations to enhance collaborating mining of knowledge at many level of abstraction. That's why data warehouse [3] has been now become special platform for data analysis & online analytical processing. Data warehousing is process of constructing & using a data warehouse. A data warehouse has been constructed by fit with group data from multiple heterogeneous sources analytical data reporting, structured and/or ad hoc queries, & decision making. Data warehousing [4] involves data cleaning, data integration, & data consolidations.

1. Introduction

The meaning Data Warehouse[1] was first coined by Bill in 1990. According to bill a data warehouse is a subject oriented, fit with group, time-variant, & non-volatile arrange of data. It data helps analysts to take informed decisions in business. An operational has database undergoes frequent changes on a daily basis on account of transactions that taken place. A business executive wants to analyze before feedback on any data such as a product, a supplier, or any consumer data, then executive would have no data available to analyze because previous data has been updated due to transactions. Using Data Warehouse Information There are decision support technologies that help utilize data able to be gotten in a data warehouse. These technologies help executives to use warehouse quickly & effectively. They could gather data, analyze it, & take decisions based on information present in warehouse. Information gathered in a warehouse could be used in any of following domains: **Tuning Production Strategies** - product strategies could be well tuned by repositioning products & managing product portfolios by comparing sales quarterly or yearly. **Client Analysis** - Client analysis is done by analyzing customer's buying preferences, buying time, etc. **Operations of Analysis** - Data warehousing [3] also helps in client relationship system, & making environmental corrections. Information also allows us to analyze business operation Integrating Heterogeneous Databases to fit with group heterogeneous databases, we have two approaches: Query-driven Approach Update-driven Approach Query-Driven Approach

This is traditional approach to integrate heterogeneous databases. This approach was used to build wrappers & integrators on top of multiple heterogeneous databases. These integrators are also known as mediators. Process of Query-Driven Approach When a query is issued to a client side, a metadata dictionary translates query into an appropriate form for individual heterogeneous sites involved. Now these queries are mapped & sent to local query processor. Results from heterogeneous sites are integrated into a global answer set. Disadvantages Query-driven[13] approach needs complex integration & filtering processes. This approach is very inefficient. It is very expensive for frequent queries. This approach is also very expensive for queries that require aggregations.

Update-Driven Approach[9]

This is an alternative to traditional approach. Today's data warehouse systems follow update-driven approach rather than traditional approach discussed earlier. In update-driven approach, information from multiple heterogeneous sources are integrated in advance & are stored in a warehouse. This information is available for direct querying & analysis. It includes tools for extracting data from multiple operational databases & external sources; for cleaning, transforming & integrating this data; for loading data into data warehouse; & for periodically refreshing warehouse to reflect updates at sources & to purge data from warehouse, perhaps onto slower archival storage. In addition to main warehouse, there may be several departmental data marts. Data in warehouse & data marts is stored & managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, & data mining tools. Finally, there is a repository for storing & managing metadata, & tools for monitoring & administering warehousing system. Warehouse may be distributed for load balancing, scalability & higher availability.

2. Literature Review

[1] Surajit Chaudhuri wrote on An Overview of Data Warehousing & OLAP Technology (Appears in ACM Sigmod Record, March 1997).[2]

Data warehousing & on-line analytical processing (OLAP) are essential elements of decision support, which has all the time more become a focus of database industry. Many commercial products & services are now available, & all of principal database management system vendors now have offerings in these areas. Decision support places some rather various necessary on database technology compared to traditional on-line transaction processing applications. This paper given an overview of data warehousing & OLAP[2] technologies, with an emphasis on their new requirements. We describe back end tools for extracting, cleaning & loading data into a data warehouse; multidimensional data models typical of OLAP; front end client tools for querying & data analysis; server extensions for efficient query processing; & tools for metadata management[10] & for managing warehouse. In addition to surveying state of art, this paper also identifies some promising

research problem, some of which are related to issue that database research community has worked on many years, but others then only just beginning to be addressed. This overview is based on a tutorial that authors presented at VLDB Conference, 1996.

[2] Manjunath T. N. wrote on Realistic Analysis of Data Warehousing & Data Mining Application in Education Domain [3]

Data-driven decision support systems, such as data warehouses could serve requirement of extraction of information more than one subject area. Data warehouses standardize data across organization so as to have a single view of information. Data warehouses could provide information required by decision makers. Developing a data warehouse for educational institute is less focused area since educational institutes are non-profit & service oriented organizations. In now day scenario where knowledge has been privatized & cut throat competition is prevailing, institutes needs to be more organized & need to take better decisions. Institute's enrollments are increasing as a result of increase in number of branches & intake. Now a day, any reputed Institute's enrollments count in to thousands. In view of these factors challenges for management are meeting diverse needs of students & facing increased complexity in academic processes. Complexity of these stimulate intellect requires continual improvements in operational strategies based on accurate, timely & consistent information. Cost of building a data warehouse is high for any educational institution as it requires data warehouse tools for building data warehouse & extracting data using data mining tools[11] from data warehouse. Present study provides an option to build data warehouse & extract useful information using data warehousing & data mining open source tools. In this paper they have explored need of data warehouse / business intelligence for an educational institute, operational data of an educational college has been used for experimentation. Study may help decision makers of educational institutes across globe for better decisions.

[3] Mr. Dishek Mankad wrote on "The Study on Data Warehouse Design & Usage"[1]

Data ware housing is a booming industry has been within many interesting research problem. Data warehouse is concentrated on only few aspects. Here we are discussing about data warehouse design & usage. Let's look at various approaches to data ware house design & usage process & steps involved. Data warehouse could be built using a top-down approach, bottom –down method or a combination of both. In this research paper we are discussing about data warehouse design process.

3. Need & Scope of Research

A data warehouse[1] is a database, which is kept separate from organization's operational database. There is no habitual updating done in a data warehouse. It possesses consolidated historical data, which helps organization to examine its business. A data warehouse helps executives to organize, understand, & use their data to take strategic decisions. Data warehouse systems help in integration of diversity of application systems. A data warehouse system helps in consolidated historical data analysis Data Warehouse[4] is separated from Operational Databases. A data warehouses is kept separate from operational databases due to following reasons: An operational database is constructed for well-known tasks & workloads such as searching particular records, indexing, etc. In contract, data warehouse queries are often complex & they present a general form of data. Operational databases support concurrent processing of multiple transactions. Concurrency control & recovery mechanisms are required for operational databases to ensure robustness & consistency of database. An operational database query allows to read & modify operations, while an OLAP[2] query needs only **read only** access of data. An able to used database maintains current data. On other hand, a data warehouse maintains historical data.

4. Proposed Work

1. Investigation of new Challenges in Data warehouse management[1].
2. Study of runaway queries problems.[12]
3. Taking corrective measurement to manage Resource
4. Taking corrective steps in order to solve Scheduling problem.

5. Challenges

We have described substantial technical challenges in developing & deploying decision support systems. While many commercial[6] products & services exist, there are still several interesting avenues for research. We would only touch on a few of these here. Data cleaning is a problem that is reminiscent of heterogeneous data equal access for all, a problem that has been studied for many years. But here emphasis is on data inconsistencies instead of schema inconsistencies thing. Data cleaning, as we indicated, is also closely related to data mining, with objective of suggesting possible inconsistencies.

6. Scope of Research

The problem of physical design of data warehouses [1] should rekindle interest in well-known problems of index selection, data partitioning & selection of materialized views. However, while go to place again these problems, it is important to recognize special role played by aggregation. Decision support systems already provide field of query optimization[13] with increasing challenges in traditional questions of selectivity estimation & cost-based algorithms that could exploit transformations without exploding search space (there are plenty of transformations, but few reliable cost estimation techniques & few smart cost-based algorithms/search strategies to exploit them). Partitioning functionality of query engine[12] between middleware (e.g., ROLAP layer) & back end server is also an interesting problem. Management of data warehouses also presents new challenges. In particular, failure & check pointing issues in load & refresh in presence of many indices & materialized views needs further research. Adaptation & use of workflow technology might help, but this needs further investigation. Detecting runaway queries & managing & scheduling resources are problems that are important but have not been well solved. Some work has been done on logical correctness of incrementally updating materialized views, but performance, scalability, & recoverability properties of these techniques have not been investigated.

References

[1] Mr. Dishek Mankad "The Study on Data Warehouse Design & Usage" International Journal of Scientific & Research Publications, Volume 3, Issue 3, March 2013 ISSN 2250- 3153

[2] Surajit Chaudhuri wrote on An Overview of Data Warehousing & OLAP Technology (Appears in ACM Sigmod Record, March 1997).

[3] Manjunath T. N. wrote on Realistic Analysis of Data Warehousing & Data Mining Application in Education Domain

[5] Kimball, R.The Data Warehouse Toolkit. John Wiley, 1996.

[6] Barclay, T., R. Barnes, J. Gray, P. Sundaresan, "Loading Databases using Dataflow Parallelism." SIGMOD Record, Vol.23, No. 4, Dec.1994.

[7] Blakeley, J.A., N. Coburn, P. Larson. "Updating Derived Relations: Detecting Irrelevant & Autonomously ComputableUpdates." ACM TODS, Vol.4, No. 3, 1989.

[8]Gupta, A., I.S. Mumick, "Maintenance of Materialized Views: Problems, Techniques, & Applications." Data Eng. Bulletin, Vol. 18, No. 2, June 1995. 9 Zhuge, Y., H. Garcia-Molina, J. Hammer, J. Widom, "View Maintenance in a Warehousing Environment, Proc. Of SIGMOD Conf., 1995.

[9] Roussopoulos, N., et al., "The Maryland ADMS Project: Views R Us." Data Eng. Bulletin, Vol. 18, No.2, June 1995.[11] O'Neil P., Quass D. "Improved Query Performance withVariant Indices", To appear in Proc. of SIGMOD Conf., 1997.

[10] O'Neil P., Graefe G. "Multi-Table Joins through BitmappedJoin Indices" SIGMOD Record, Sep 1995.

[11] Harinarayan V., Rajaraman A., Ullman J.D. " Implementing Data Cubes Efficiently" Proc. of SIGMOD Conf., 1996.

[12] Chaudhuri S., Krishnamurthy R., Potamianos S., Shim K."Optimizing Queries with Materialized Views" Intl.Conference on Data Engineering, 1995.

[13] Levy A., Mendelzon A., Sagiv Y. "Answering Queries Using Views" Proc. of PODS, 1995. 16 Yang H.Z., Larson P.A. "Query Transformations for PSJ Queries", Proc. of VLDB, 1987

